(19) Japan Patent Office (JP)
(11) Japanese Unexamined Patent Application Publication
(12) Public Patent Information (A) S59 – 44648

(51) Int. Cl. [3]	ID. No.	Internal Filing. No.
G 01 N 27/26	A	7363 – 2G
33/48	Z	8305 – 2G
G 01 T 1/08		8105 – 2G
G 21 H 5/02		8204 – 2G

(43) Publication Date: March 13th, 1984

Number of Inventions: 1
Examination Request: Not Filed
(Total Pages: 6)

---

(54) Title of the Invention: METHOD OF DETERMINING THE BASE SEQUENCES OF DIOXYRIBONUCLEIC ACIDS

(21) Application No.: S57 – 155550
(22) Application Filed: September 7th, 1982

(72) Inventor: Takashi SATOH
829 – 6 Jurokken Kumagaya City
(71) Applicant: Mimata Seisakusho Co., Ltd.
1 – 12 – 13 Uchikanda, Chiyoda-ku, Tokyo

(74) Representative: Kyozo YUASA, Patent Attorney (and 4 others)

Specification

1. Title of the Invention: METHOD OF DETERMINING THE BASE SEQUENCES OF DIOXYRIBONUCLEIC ACIDS

2. Scope of Patent Claims

A method for determining the base sequences of DNA fragments with the following characteristics. One end of DNA segments were labeled with a radioactive element and subjected to base-specific cleaving. The DNA fragment samples produced by that cleavage underwent electrophoresis and then many of the labeled DNA fragments were detected using autoradiography. The base sequences of DNA segments produced in this way were determined using methods in which base-specific cleaving was used on 4 types of DNA fragment samples, which were then mixed to form a reference sample. These were lined up alternately and DNA fragment samples that had been subjected to base-specific cleaving were placed in between them. These were subjected to electrophoresis and, based on the results of electrophoresis on the reference samples on both sides of the samples, we predicted the positions of zones with base counts that were off by one in the DNA fragment line in one of the samples that had been subjected to base-specific cleaving. We determined whether or not the zone corresponding to that position existed in the sample line that had been subjected to base-specific cleaving and we determined the phoresis position of many of the DNA fragments in that sample. We performed the same sort of procedure on many DNA fragments in the other samples that had undergone base-specific cleaving, which allowed us to determine the positions of many of the DNA fragments in the other samples that had undergone base-specific cleaving. This allowed us to determine the order of each of the fragments that had undergone base-specific cleaving.

3. Detailed Description of the Invention

This invention pertains to methods of determining the base sequences of dioxyribonucleic acid (DNA) using electrophoresis.

In this specification, the term "DNA segments" refers to DNA composed of several hundred base pairs that are used as samples for determining the base sequences of the DNA using the Maxam-Gilbert or other methods and then using a restriction enzyme to cleave gigantic ring or linear DNA that exists in organisms. The term "DNA fragments" refers to DNA segments that have been cleaved to individual lengths using the Maxam-Gilbert or other

methods and cleaved at a specific base with particular methods. The term "DNA fragment sample" refers to a mixture of the individual DNA fragments of all lengths that result when a DNA segment has been cleaved at a specific base.

Recently, with the development of genetic engineering, there has been a rapidly growing need to determine the base sequences of DNA. in the field of molecular biology. As a matter of fact, a number of methods, starting with the Maxam-Gilbert Method, have been developed for determining base sequences.

However, with the Maxam-Gilbert Method, DNA that has been labeled with the radioactive element [^{32}P] is cleaved in a base specific way using chemicals. That is, samples of DNA fragments that have been cleaved at specific sites (T (thymine), C (cytosine), G (guanine) and A (adenine)) are lined up from left to right in specific positions on polyacrylamide gel. Then, using polyacrylamide gel electrophoresis and they are separated using differences in length of a single base and subjected to autoradiography. (Figure 1). The position of the zones that appear in the autoradiogram correspond to one of the positions that the respective bases occupy in the ONA and, the farther away the zone has moved, the closer the specifically cleaved base is to the end of the DNA. Actually, the method determines the base sequences of the DNA by starting with the distant zone that moved the farthest away and determining in order, which ONA fragment was cleaved with which base.

It is said that this method is able to determining around 200 ~ 250 base sequences with a 20 x 40 cm of polyacrylamide gel, however, to do so requires measuring the distances traveled (the position of the zones) by each DNA fragment cleaved specifically at its respective base position in the autoradiogram and to compare them with the distances traveled by the other DNA fragments. The distances traveled by the DNA fragments at this time are inversely proportional to the logarithm of the size of the fragment so in the areas where the DNA fragments are comparatively large, the difference in the amount of movement in each of the zones will be extremely small, demanding highly precise position measurements.

However, at present, the automatic reading and analysis of the zones on the autoradiogram using the Maxam-Gilbert Method has not yet been systematized and the determinations are made by having people compare the positions of the zones using just their eyes.

However, there is one significant problem in building an automatic DNA base sequence reading and analysis system. That is, the lack of uniformity in the polymerization of the polyacrylamide gel or the lack of uniformity in the electric field that is used in electrophoresis. For these reasons, the zones of the DNA fragments that are separated using phoresis are not at precisely right angles in relation to the phoresis direction. They sometimes bend or become distorted (Figure 1). However, it is difficult to take this distortion out of electrophoresis. To expand an example of this distortion, the G zone in Figure 2 is at a right angle to the direction of phoresis, but there was less phoresis on the left side than on the right side and it rises to the left. Furthermore, the degree of inclination for the T zone is much worse than for C. When this distortion was observed and the order of the zones was being determined visually, it was first thought that the T zone was an extended line in the C direction and a decision was made as to whether the extended line was above or below the C zone. Next, the relative positions of the C and G zones were determined in the same manner. (In this case, the local base sequence was TCG.) The simplest method for determining the positions of these zones is to measure the position of the center point of each of the zones (y coordinate), compare that value and determine the order of the zones. However, in the example shown in Figure 2, if the center point of the respective TCG zones are yT, yC and yG, then yT > yC > yG. Considering the method of taking the coordinate axis with these measurements, then the smaller y values would indicate greater phoretic distance. In this case, the local base sequence would be GCT, which is clearly different than when determining the order of the zones visually.

In this way, in devices that automatically read DNA base sequences, it is meaningless simply to measure

the center point of each zone in the autoradiogram accurately. In particular, in areas where the DNA fragments are large, some sort of method for correcting the zone distortion must be developed that corresponds to determining the relative positions of the zones visually. With this method, the way to line up the samples for the polyacrylamide gel electrophoresis was to line up sequentially from left to right, the DNA fragment samples that had been cleaved specifically at their bases T, C, G and A (Figure 1). As a result of using this sort of sample arrangement for electrophoresis, it was necessary to correct any distortion or bending of the zones that was caused by incomplete electrophoresis for the determination of the positions of the DNA fragment zones on the autoradiogram. Not doing so would have caused imprecise readings as discussed in the section "Shortcoming of Conventional Technology." The first method for making this sort of correction is a method that uses a computer to do the exactly the same correction that is done when performing the readings visually. That is, in Figure 2, the coordinates of the left edge and the right edge of the T zone are calculated, and this value is used to calculate the extended line in the C direction of this zone. The positions of the center point of the C zone and this extended line are compared and it can be determined which zone went farther during the phoresis.

This process is repeated for all of the zones T, C, G and A, which makes it possible to determine the order of each zone. However, the disadvantage is that doing so not only requires a computer that is capable of doing a considerable number of calculations and that has a large memory capacity, but the number of measuring positions increases, so the amount of error also increases.

In this invention, as shown in Figure 2, the samples are lined up as follows. DNA fragments are cleaved specifically to each base. All of the DNA fragments that have been subjected to the four types of base-specific cleaving on both ends are mixed together. That mix is used as a reference to line up the samples, which eliminates the correction of the error cause by zone position distortion.

This invention pertains to methods of determining the base sequences of DNA segments with the following features. One end of a DNA fragment is labeled with a radioactive element made up of [^{32}P] and that fragment is subjected to base-specific cleaving. The DNA fragment sample that results from the cleaving then undergoes electrophoresis. Many of the DNA fragments that were labeled are detected using autoradiography. In methods of determining the base sequences of DNA segments that are made up of such fragments, the 4 types of DNA fragment samples that were subjected to each of the base-specific cleavages are mixed together and lined up alternately with reference samples. DNA fragment samples that were subjected to base-specific cleaving are placed between them. Electrophoresis is performed and a prediction is made of the position to which the zones will move during electrophoresis. (The base count of those zones differs by one along the line of DNA fragments in a sample that had been subjected to base-specific cleaving.) A decision is then made regarding whether or not the zone corresponding to that position, is in the line of the sample that was subjected to base-specific cleaving. The phoresis positions of many of the DNA fragments in that sample will be determined and many of the DNA fragments in the other samples that were also subject to base-specific cleaving will undergo the same sort of process. This is how the base-specific cleaving was carried out. This will determine many of the DNA fragment positions of the samples with that value, which will allow the order of each of the fragments that had been subjected to base-specific cleaving to be determined.

The phoretic positions of many of the DNA fragments contained in a DNA fragment sample and the phoretic positions of DNA fragments contained in other samples can be measured at the same time with 4 DNA fragment samples lined up between 5 comparison samples shown in Figure 3 or they can be measured separately. As shown in Figure 4, for nearly all of the zone distortion and bending caused by incomplete electrophoresis, those closest to the side surface of the gel have the shortest phoresis distance, so the left side of the gel inclines upward to the left and a zone that goes up on the right forms the right side of the gel. For this reason, the positions of the nth zones (y'n and y"n) in comparison 1 and comparison 2 are measured accurately and the average y coordinate values are calculated ($yTn = (y'n + y"n)/2$). This yTn indicates the position that comes about as a result of the phoresis of the DNA fragments made up of n nucleotides in the T line. Next, it is determined whether or not a zone corresponding to this position exists in the T line. In the example shown in Figure 4, the zone in

question does not exist. In the example using the T line in Figure 4, the Reference 1, T and Reference 2 zones are inclined to nearly the same degree with respect to the phoresis direction. It is also mathematically clear that the positions to which the DNA fragments, composed of "n" nucleotides at the T line, should come to by phoresis as calculated using the above method are exactly the same as those positions to which the fragments of this size come to on the conventional T line by phoresis. However, as seen in the Reference 2, C and the Reference 3, in Figure 4, when the inclination differs slightly in each of the respective zones (experience shows that the size of the inclination is Reference 2 > C > Reference 3), the phoresis positions calculated along the C line are ycn = (y"n + y'"n)/2 are very slightly out of alignment with the actual positions y cm where the DNA fragments made up of "n" nucleotides come to via phoresis. However, the degree of this misalignment is far smaller than the difference between itself and the position to which the neighboring DNA fragments of differing lengths come to by phoresis. If these can be kept within a certain threshold value (if, for example, the differences between contiguous zones is kept to within 30%), then those zones could be judged to exist. In this way, the respective positions of the zones of the DNA fragments that underwent base-specific cleaving directly are not compared. References must be placed on both sides of the sample and the phoresis positions in the sample lines of the DNA fragments that are one unit different in length in comparison with the references are predicted. By determining whether or not a DNA fragment zone exists at that position, it is possible to correct effectively and easily, the distortion and bending that accompany incomplete electrophoresis.

An example of the procedure for analyzing the data of this invention is described below.

Using a computer that is connected to a densitometer the following measurements and calculations are performed.

4 – 1 Check for Existence of a Chemically Modified Base

In Figure 3, the line for Reference 1 is scanned in the direction of the arrow and the positions of the center points of the respective zones are read in and stored. These are subject to chemical modification in this DNA and, in order to determine whether or not there are bases that have not been cleaved, a check is made to see if the following formula holds among the positions yn, yn + 1 and yn + 2 of the three continuous zones n, n + 1 and n + 2.

If this formula Yn + 2 – Yn + 1 < Yn + 1 – Yn (n = 1, 2, 3...) does not hold, it means that there is a chemically modified base between the $n + 1^{th}$ and the $n + 2^{th}$ nucleotides. The position of this base is stored and corrected against the base count in the subsequent calculations.

4 – 2 Determining the T Position in the DNA

Using the densitometer, the center points (y'n, ytn, y"n, where n = 1, 2, 3...) of the zones of the lines of Reference 1, T and Reference 2 in Figure 2 are measured and stored in a computer. The averages (yTn, where n =1, 2, 3...) of the positions of the zones with numbers corresponding to Reference 1 and Reference 2 are taken. Next, the formula below is applied to the positions (yTm, where n(?) = 1, 2, 3...) of the center points of the T zone that were stored, to find zones corresponding to yTn.

1 yTm – yTn 1 < 0.31 yFn – yTn + 11

(where n = 1, 2, 3..., m = 1, 2, 3...)

If such zones exist, they are stored as having a T where the n values are at those times. The n value indicates that the n^{th} nucleotide from the end of the DNA is T.

4 – 3 Determining the Position of Other Bases in the DNA

Using the same methods as in 4 – 2, the C position was calculated using Reference 2 and Reference 3 in Figure 3, the G position from Reference 3 and Reference 4 and the A position was determined using Reference 4 and Reference 5. These values were then stored.

4 – 4 Determining the DNA Base Sequences

From among the stored base position (number) data, the bases corresponding to the first, second... were called in sequence and those numbers were printed out.

4 – 5 Verifying the Base Sequences Using a Complementary DNA Strand

It is possible to determine DNA base sequences using the methods described above, but due to experimental error, it is possible that errors occurred in the data analysis. In order to increase the certainty, as shown in Figure 5, the DNA for which the base sequences are to be determined and a sample of a complementary DNA

strand are lined up on a polyacrylamide gel for phoresis. Using the above methods of analysis, the two DNA base sequences produced are examined to see if they are complementary or not, taking the cleavage positions and the size of each DNA sample into account. That is, a check is performed to see if the A – T and G – C airs have formed in the base sequences. If they are not complementary, then that fact is disclosed.

4. Brief Description of the Drawings

Figure 1 is a schematic diagram showing how the samples are lined up on the polyacrylamide gel according to the Maxam-Gilbert Method. T, C, G and A are the DNA fragment samples that each underwent base-specific cleaving.

Figure 2 is a schematic diagram showing the imprecision of measurements when there is distortion or bending in the zone. The center points of the zones yT, yC and yG are shown.

Figure 3 is a schematic diagram showing how the samples are lined up on the polyacrylamide gel in this invention.

Figure 4 is a schematic diagram of the correction of zone distortion and bending in this invention.

Figure 5 is a schematic diagram showing how the samples are lined up on the polyacrylamide gel when checking the complementariness of the DNA.

Patent Applicant: Mimata Seisakusho Co., Ltd.

Representative: Kyozo YUASA, Patent Attorney (and 5 others) [illegible stamp]

Figure 1

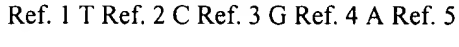
Initial Phoresis Position

Electrophoresis

Figure 2

Electrophoresis

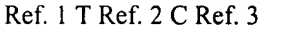
Figure 3



Initial Phoresis Position

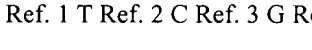
Electrophoresis

Figure 4



Electrophoresis

Figure 5



Complementary Strand